

Algorithm Design for Resource Management

Of Time Critical Applications in Edge-Cloud Architecture

Student: Shaan Aryaman Supervisor: Assoc. Prof. Arvind Easwaran

Motivation and Problem Formulation:

Traditional cloud servers are situated far away from end-devices. This increases transmission time and makes it impossible to service time critical applications. With the recent advent of low latency 5G wireless technology, edge computing is gaining popularity. Edge servers or Access points (APs) are small servers situated next to end devices. In our resource management problem formulation each task has an associated profit value and deadline by which it must be processed. Each task must be mapped to an AP to be offloaded to the edge-cloud architecture and then subsequently mapped to a cloud server to be processed. Our problem accounts for two types of resources: (i) bandwidth (ii) compute that must be allocated by an AP and a cloud server respectively to a task such that its deadline is met. It is interesting to note that the multi-dimensional multiple knapsack problem is a special case of our problem.

Algorithms Devised:

1) Heuristic Algorithm: Zero Slack Greedy (ZSG) algorithm. This algorithm enumerates all task, AP, cloud server combinations and based on heuristic assignment chooses which combinations will be used to allocate tasks to appropriate AP and cloud servers and with what amount of resource. The algorithm is very time-efficient and lightweight.



2) Approximation Algorithm: Devised a novel Polynomial Time Approximation Scheme (PTAS) for the discretised version of the problem and when the total number of servers is a constant.

Results:

For evaluating ZSG we created 12,000 task-sets with varying numbers of tasks, compute intensive tasks, bandwidth intensive tasks. APs and cloud servers remained constant across task-sets. All figures show the Profit Gain Ratio (PGR) = (total profit of provisioned tasks)/(total profit of all tasks in task-set) in the y-axis. At-least 70% PGR achieved by ZSG in all scenarios on average.



These results are from our paper: C. Gao, A. Shaan, A. Easwaran. "Deadline-constrained Multi-resource Task Mapping and Allocation for Edge-Cloud Systems," in IEEE Globecom, Rio de Janeiro, Brazil, 2022. Public version at : https://arxiv.org/abs/2206.05950

www.scse.ntu.edu.sg